Probabilités 2 - Variables aléatoires sur un univers fini

Dans tout ce chapitre, $(\Omega, \mathbb{P})$ désigne un espace probabilisé fini.

1 Notion de variable aléatoire

Définition. *On appelle* ***variable aléatoire*** *(v.a.) toute fonction $X : \Omega \to E$, où E est un ensemble quelconque.*

Lorsque $E \subset \mathbb{R}$, on dit que X est une v.a. réelle (v.a.r.).

Notations.

- *L'ensemble des valeurs possibles pour X est fini. En effet, Ω est fini ; mettons, $\Omega := \{\omega_1, \dots, \omega_k\}$. Donc, pour tout $\omega \in \Omega$, on a $X(\omega) \in \{X(\omega_1), \dots, X(\omega_k)\}$.*
  Si $x_1, \dots, x_n$ désigne les valeurs possibles de X, l'ensemble $\{x_1, \dots, x_n\}$ est noté $X(\Omega)$ (notation logique, puisqu'il s'agit bien de l'image directe de Ω par X).
- *Si $A \subset E$, l'évènement $\{\omega \in \Omega \mid X(\omega) \in A\}$ est simplement noté $(X \in A)$, ou encore $\{X \in A\}$. Pour une v.a.r., on notera par exemple $(X \leqslant 3)$ l'évènement $\{\omega \in \Omega \mid X(\omega) \leqslant 3\}$.*

Théorème/définition. *Avec les notations précédentes, l'application*

$$\mathbb{P}_X : \begin{array}{ccc} \mathcal{P}(X(\Omega)) & \longrightarrow & [0,1] \\ A & \longmapsto & \mathbb{P}(X \in A) \end{array}$$

est une probabilité sur $X(\Omega)$, appelée ***loi de probabilité de X*** *(ou simplement loi de X).*

Démonstration. Simples vérifications sur le fait que $\mathbb{P}_X(X(\Omega)) = 1$ et $\mathbb{P}_X(C \cup D) = \mathbb{P}_X(C) + \mathbb{P}_X(D)$ si $C \cap D = \varnothing$. □

Remarque fondamentale : Le fait que Ω soit fini a pour conséquence que la loi de X est entièrement caractérisée par :

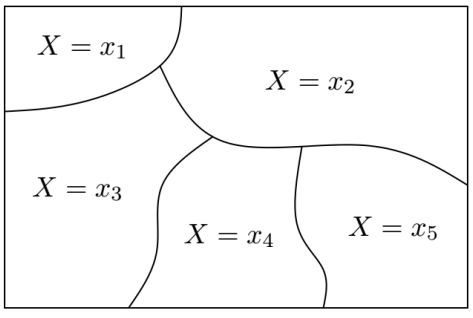
1. l'ensemble $X(\Omega) = \{x_1, \dots, x_n\}$ des valeurs qu'elle prend ;
2. la distribution de probabilités $p_1, \dots, p_n$, où pour tout $i \in [\![1; n]\!]$, $p_i := \mathbb{P}(X = x_i)$.

Exemple : Si $A = \{x_2, x_5, x_7\}$, alors :

$$\mathbb{P}(X \in A) = \mathbb{P}\left[(X = x_2) \cup (X = x_5) \cup (X = x_7)\right] = p_2 + p_5 + p_7$$

On peut ainsi obtenir $\mathbb{P}(X \in A)$ pour tout $A \subset X(\Omega)$.

La famille d'évènements $(X = x_i)_{i \in [\![1;n]\!]}$ est un système complet d'évènements (SCE) de Ω, et on a : $\sum_{i=1}^{n} p_i = 1$.



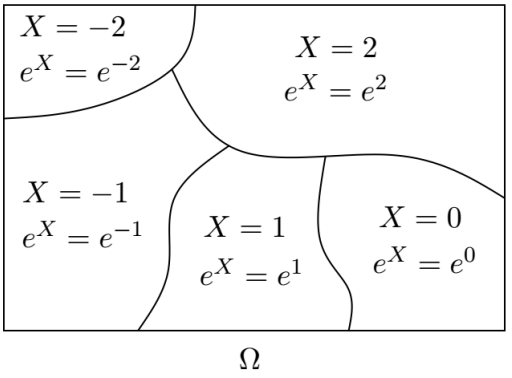
Prenons l'exemple d'une v.a.r. qui prend cinq valeurs :

Valeurs ($X(\Omega)$)	-2	-1	0	1	2
Probas	p_1	p_2	p_3	p_4	p_5

On cherche à déterminer la loi de $f(X)$, où f est une fonction.

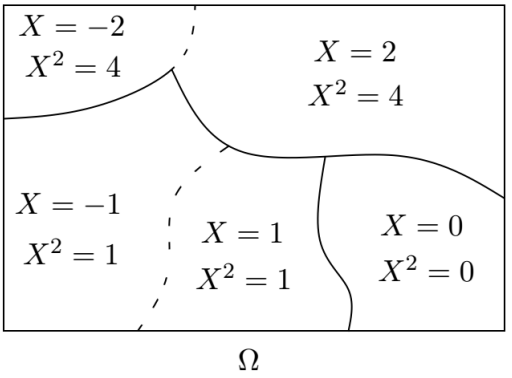
▶ **Exemple n° 1 :** Soit $f : t \mapsto e^t$. Lorsque $X = x_i$, e^X vaut e^{x_i}. On en déduit le tableau de valeurs suivant :

Probas	p_1	p_2	p_3	p_4	p_5
Valeurs de X	-2	-1	0	1	2
Valeurs de e^X	e^{-2}	e^{-1}	e^0	e^1	e^2



▶ **Exemple n° 2 :** Soit $f : t \mapsto t^2$.
Lorsque $X = x_i$, X^2 vaut x_i^2. On en déduit le tableau de valeurs suivant :

Probas	p_1	p_2	p_3	p_4	p_5
Valeurs de X	-2	-1	0	1	2
Valeurs de X^2	4	1	0	1	4



On remarque dans ce deuxième exemple que l'ensemble $X^2(\Omega)$ est simplifié, car $X^2(\Omega) = \{0, 1, 4\}$. On a de plus

Valeurs de X^2	0	1	4
Probas	p_3	$p_2 + p_4$	$p_1 + p_5$

Notation. *On note $X \sim Y$ pour signifier que X et Y ont la même loi, c-à-d que $\mathbb{P}_X = \mathbb{P}_Y$.*

Remarque. *Si $X \sim Y$, alors pour toute fonction f, $f(X) \sim f(Y)$.*

2 Lois usuelles

2.1 Loi uniforme sur $\{x_1, \ldots, x_n\}$

Définition. *Soit X une v.a. et $x_1, \ldots, x_n$ des éléments deux à deux distincts d'un ensemble E.*
*On dit que X suit la **loi uniforme** sur $\{x_1, \ldots, x_n\}$, et on note $X \sim \mathcal{U}(\{x_1, \ldots, x_n\})$ si :*

- $X(\Omega) = \{x_1, \ldots, x_n\}$
- $\forall i \in [\![1; n]\!]$, $\mathbb{P}(X = x_i) = \dfrac{1}{n}$

2.2 Loi de Bernoulli $\mathcal{B}(p)$

Définition. *On dit qu'une v.a. X suit la* ***loi de Bernoulli*** *de paramètre $p \in [0,1]$, et on note $X \sim \mathcal{B}(p)$ si :*

- $X(\Omega) = \{0,1\}$
- $\mathbb{P}(X=1) = p$
- *(et donc)* $\mathbb{P}(X=0) = 1-p$

Intérêt : Il s'agit du modèle élémentaire pour le succès/l'échec d'une expérience aléatoire.

Exemple : Dans une urne, on place 10 boules, numérotées de 1 à 10.

On tire une boule, et on note X le numéro de la boule tirée. On note $Y = \mathbb{1}_{(X \geqslant 7)} = \begin{cases} 1 \text{ si } X \geqslant 7 \\ 0 \text{ sinon} \end{cases}$ (Y est appelée fonction indicatrice de l'évènement $(X \geqslant 7)$).

Alors, Y suit la loi de Bernoulli de paramètre $p = \dfrac{4}{10} = \dfrac{2}{5}$.

2.3 Loi binomiale $\mathcal{B}(n,p)$

Remarque fondamentale : Soit $X_1, \dots, X_n$ des v.a.r. suivant $\mathcal{B}(p)$ (où $p \in [0;1]$ fixé) indépendantes, au sens où les n évènements de succès $A_i := (X_i = 1)$ sont indépendants.

Autrement dit, on répète n fois une expérience qui a pour probabilité p de réussir.

On pose $S :=$ le nombre de succès sur les n expériences.

Alors S est une v.a.r., et $S(\Omega) = \{\underbrace{0}_{\substack{\text{aucun} \\ \text{succès}}}, \ 1 \ , \ 2 \ , \ \dots \ , \ \underbrace{n}_{\substack{\text{que des} \\ \text{succès}}}\}$.

- Premièrement : S est la **somme** des X_i.
  En effet, $\underbrace{X_1 + X_2 + \dots + X_n}_{\text{somme de 0 et de 1}} = \big[1 \times (\text{nb de succès})\big] + \big[0 \times (\text{nb d'échecs})\big] = S$

- Deuxièmement : quelle est la loi de S ? Soit $k \in [\![0;n]\!]$; calculons $\mathbb{P}(S=k)$.
  On regarde pour cela les séquences $(X_1, \dots, X_n)$ qui ont exactement k succès.

  Exemples :
  - $\underbrace{1\,1\,\dots\,1}_{k \text{ fois}}\underbrace{0\,0\,\dots\,0}_{n-k \text{ fois}}$: cette séquence a pour probabilité (par indépendance) :
  $$\underbrace{p \times p \times \dots \times p}_{k \text{ fois}} \times \underbrace{(1-p) \times (1-p) \times \dots \times (1-p)}_{n-k \text{ fois}} = p^k (1-p)^{n-k}$$
  - $\underbrace{0\,0\,\dots\,0}_{n-k \text{ fois}}\underbrace{1\,1\,\dots\,1}_{k \text{ fois}}$: cette séquence a pour probabilité (par indépendance) :
  $$\underbrace{(1-p) \times (1-p) \times \dots \times (1-p)}_{n-k \text{ fois}} \times \underbrace{p \times p \times \dots \times p}_{k \text{ fois}} = (1-p)^{n-k} p^k$$

  Ces exemples illustrent le fait que le nombre de séquences favorables est $\binom{n}{k}$, car il s'agit du nombre de façons de placer les 1 dans n emplacements.

  La probabilité de chaque séquence étant $p^k(1-p)^{n-k}$, on en déduit que :

$$\boxed{\mathbb{P}(S=k) = \binom{n}{k} p^k (1-p)^{n-k}}$$

Théorème/définition. *On dit qu'une v.a.r. X suit la loi binomiale de paramètres $n \in \mathbb{N}^*$ et $p \in [0;1]$, et on note $X \sim \mathcal{B}(n,p)$, si :*

$$\begin{cases} X(\Omega) = [\![0;n]\!] \\ \forall k \in [\![0;n]\!],\ \mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \end{cases}$$

De plus,

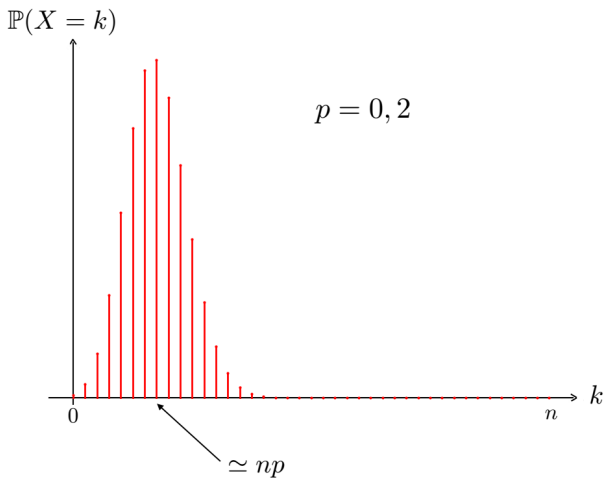
- *Si $X_1, \dots, X_n$ sont n v.a.r. indépendantes, toutes de loi $\mathcal{B}(p)$, alors la v.a.r. $S := X_1 + \cdots + X_n$ suit la loi $\mathcal{B}(n,p)$.*
- *Si $A_1, \dots, A_n$ sont n évènements indépendants, tous de probabilité p, alors si S est « le nombre de succès », c-à-d le nombre d'évènements A_i qui se réalisent, on a : $S \sim \mathcal{B}(n,p)$.*

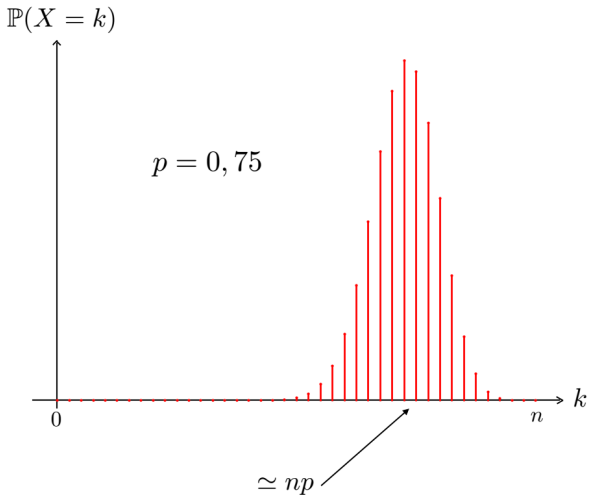
Remarques :

1. $\displaystyle\sum_{k=0}^{n} \mathbb{P}(X=k) = 1$ se retrouve avec le binôme de Newton :

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1^n = 1$$

2. Quelle est l'allure de ces probas (à n fixé) ?





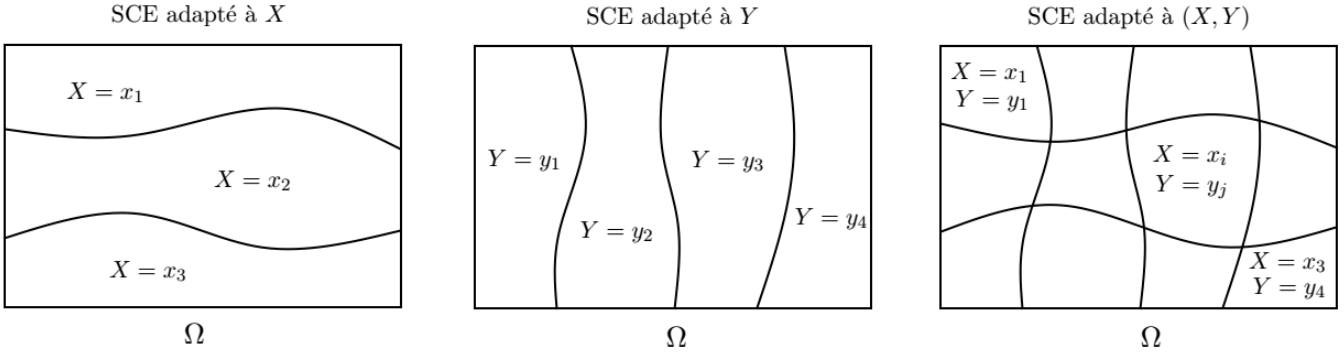
3 Couples de variables aléatoires

Introduction : Si

- X est une v.a. à valeurs dans $X(\Omega) = \{x_1, \dots, x_n\}$
- Y est une v.a. à valeurs dans $Y(\Omega) = \{y_1, \dots, y_p\}$

alors le couple (X, Y) est une v.a. dont les valeurs possibles sont les couples (x_i, y_j), où $i \in [\![1; n]\!]$, et $j \in [\![1; p]\!]$. Autrement dit :

$$(X, Y)(\Omega) = X(\Omega) \times Y(\Omega) \qquad \text{(produit cartésien)}$$



La famille d'évènements élémentaires $((X = x_i) \cap (Y = y_j))_{\substack{1 \leqslant i \leqslant n \\ 1 \leqslant j \leqslant p}}$ est un SCE adapté à (X, Y).

Définition. *La loi du couple (X, Y), ou **loi conjointe**, est la donnée d'une famille $((x_i, y_j), p_{i,j})_{\substack{1 \leqslant i \leqslant n, \\ 1 \leqslant j \leqslant p}}$, où :*

- $X(\Omega) = \{x_1, \dots, x_n\}$
- $Y(\Omega) = \{y_1, \dots, y_p\}$
- $\forall (i, j) \in [\![1; n]\!] \times [\![1; p]\!]$, $p_{i,j} := \mathbb{P}((X = x_i) \cap (Y = y_j))$ *(aussi notée : $\mathbb{P}(X = x_i,\ Y = y_j)$)*

Remarque. *On a $\displaystyle\sum_{i=1}^{n} \sum_{j=1}^{p} p_{i,j} = 1$. Autrement dit, les $(p_{i,j})_{\substack{1 \leqslant i \leqslant n \\ 1 \leqslant j \leqslant p}}$ sont une distribution de probabilités.*

Exemple n° 1 : On jette deux dés, et on note X, resp. Y, le résultat du dé 1, resp. du dé 2.

- ▶ **Loi de X** : on a $X(\Omega) = [\![1; 6]\!]$, et $\forall i \in [\![1; 6]\!]$, $\mathbb{P}(X = i) = \dfrac{1}{6}$. Ainsi, $X \sim \mathcal{U}([\![1; 6]\!])$.
- ▶ **Loi de Y** : de même, $Y \sim \mathcal{U}([\![1; 6]\!])$.
- ▶ **Loi de (X, Y)** : on a $(X, Y)(\Omega) = [\![1; 6]\!] \times [\![1; 6]\!]$, et $\forall (i, j) \in [\![1; 6]\!]^2$,
  $p_{i,j} = \mathbb{P}((X = i) \cap (Y = j)) = \dfrac{1}{6} \times \dfrac{1}{6} = \dfrac{1}{36}$, par indépendance des évènements $(X = i)$ et $(Y = j)$.

  On en déduit que $(X, Y) \sim \mathcal{U}\left([\![1; 6]\!]^2\right)$.

i \ j	1	2	3	4	5	6
1	1/36	1/36	…	…	…	1/36
2	1/36	…				1/36
3	⋮					⋮
4	⋮					⋮
5	⋮					⋮
6	1/36	…	…	…	…	1/36

Exemple n° 2 : On lance un dé, et on note X le résultat (c-à-d le numéro de la face du haut), et Y le numéro de la face opposée (c-à-d celle qui pointe vers le sol).

- **Loi de X** : on a de même que précédemment, $X \sim \mathcal{U}([\![1;6]\!])$.
- **Loi de Y** : $Y \sim \mathcal{U}([\![1;6]\!])$, car pour tout $k \in [\![1;6]\!]$, $\mathbb{P}(Y=k) = \mathbb{P}(X = 7-k) = \dfrac{1}{6}$ (dans un dé, la somme des faces opposées fait toujours 7).
- **Loi de (X,Y)** : la loi est donnée par le tableau donnant les valeurs des $p_{i,j} := \mathbb{P}((X=i) \cap (Y=j))$, pour tout $1 \leqslant i \leqslant 6$ et pour tout $1 \leqslant j \leqslant 6$.

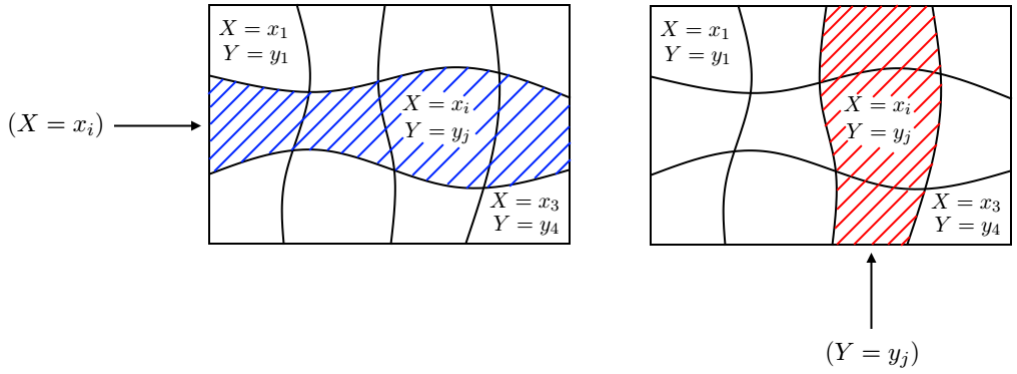
i \ j	1	2	3	4	5	6
1	0	…	…	…	0	1/6
2	⋮			⋰	1/6	0
3	⋮		⋰	1/6	⋰	⋮
4	⋮	⋰	1/6	⋰		⋮
5	0	1/6	⋰			⋮
6	1/6	0	…	…	…	0

Définition. *La loi de X et la loi de Y sont appelées les **lois marginales** du couple (X,Y).*

⚠ La connaissance des lois marginales ne suffit pas pour déterminer la loi du couple (*cf.* exemples précédents).

Théorème. *Dans les notations précédentes, on a :*

$$\mathbb{P}(X=x_i) = \sum_{j=1}^{p} p_{i,j} \qquad \text{et} \qquad \mathbb{P}(Y=y_j) = \sum_{i=1}^{n} p_{i,j}$$



Démonstration. $(Y=y_j)_{1\leqslant j\leqslant p}$ est un SCE, donc : $(X=x_i) = \bigcup_{j=1}^{p} [(X=x_i) \cap (Y=y_j)]$.

La réunion portant sur des évènements disjoints, on en déduit que

$$\mathbb{P}(X=x_i) = \sum_{j=1}^{p} \mathbb{P}\left[(X=x_i) \cap (Y=y_j)\right]$$

□

Remarque. *Il s'agit d'une formule des probabilités totales !*

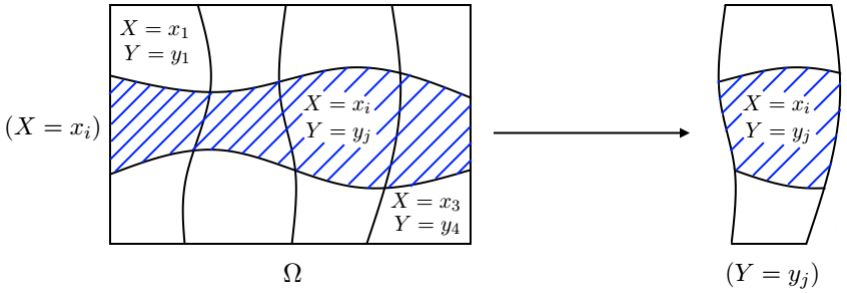
Corollaire. *On a*

$$\mathbb{P}(X = x_i) = \sum_{j=1}^{p} \mathbb{P}(Y = y_j) \times \mathbb{P}_{(Y=y_j)}(X = x_i)$$

À retenir : « on conditionne par rapport à Y ».

Définition. *Si l'on fixe un $j \in [\![1; p]\!]$, la* ***loi conditionnelle de X sachant $(Y = y_j)$*** *est la distribution de probabilités :*

$$\left(\mathbb{P}_{(Y=y_j)}(X = x_i)\right)_{1 \leqslant i \leqslant n}$$



De façon plus générale, la **loi conditionnelle de X sachant un évènement A** est la distribution de probabilités :

$$\left(\mathbb{P}_A(X = x_i)\right)_{1 \leqslant i \leqslant n}$$

Cf exo 1

Théorème/définition.

- *Un n-uplet de v.a. est une v.a. du type :*

$$Z : \begin{array}{ccc} \Omega & \to & E_1 \times \cdots \times E_n \\ \omega & \mapsto & (X_1(\omega), \ldots, X_n(\omega)) \end{array}$$

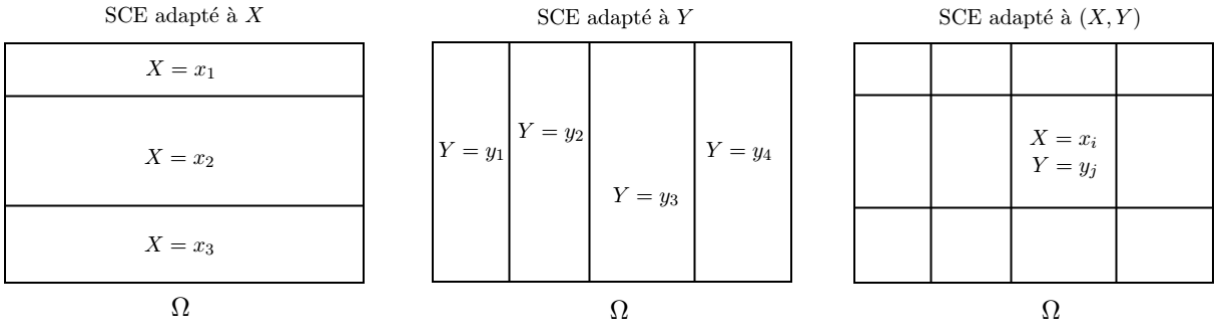
  où $E_1, \ldots, E_n$ sont des ensembles.
- *La loi de X_1 sur E_1, ..., la loi de X_n sur E_n sont appelées* ***lois marginales de Z****. Elles ne caractérisent pas la loi de Z.*
- *En revanche, les $\mathbb{P}((X_1 = x_1) \cap \ldots \cap (X_n = x_n))$, où $x_1 \in E_1$, ..., $x_n \in E_n$, forment une distribution de probabilités qui caractérise la loi de Z.*

4 Variables aléatoires indépendantes

Définition. *Soit X, Y deux v.a. et $X(\Omega) = \{x_1, \dots, x_n\}$, $Y(\Omega) = \{y_1, \dots, y_p\}$.*

*On dit que X et Y sont **indépendantes**, et on note $X \perp\!\!\!\perp Y$, si pour tout $i \in [\![1; n]\!]$ et pour tout $j \in [\![1; p]\!]$, les évènements $(X = x_i)$ et $(Y = y_j)$ sont indépendants, c-à-d si l'on a*

$$\underbrace{\mathbb{P}\left((X = x_i) \cap (Y = y_j)\right)}_{p_{i,j}} = \mathbb{P}(X = x_i) \times \mathbb{P}(Y = y_j)$$



« **Théorème** ». *Si $X \perp\!\!\!\perp Y$, alors n'importe quoi qui ne dépend que de X est indépendant de n'importe quoi qui ne dépend que de Y.*

Théorème. *Si $X \perp\!\!\!\perp Y$, alors :*

1. *Pour tout $A \subset X(\Omega)$ et pour tout $B \subset Y(\Omega)$, les évènements $(X \in A)$ et $(Y \in B)$ sont indépendants, et donc :*

   $$\mathbb{P}\left((X \in A) \cap (Y \in B)\right) = \mathbb{P}(X \in A) \times \mathbb{P}(Y \in B)$$

   Exemple : si $X \perp\!\!\!\perp Y$, alors $(-2 \leqslant X < 10) \perp\!\!\!\perp (Y^2 + 2Y - 3 = 0)$.

2. *Pour toutes fonctions f et g, les v.a. $f(X)$ et $g(Y)$ sont indépendantes.*

   Exemple : si $X \perp\!\!\!\perp Y$, alors $(e^{2X} + X) \perp\!\!\!\perp \ln(1 + Y^2)$.

Définition. *Les n v.a. $X_1, \dots, X_n$ sont dites **indépendantes** lorsque pour toutes valeurs $x_1, \dots, x_n$ prises par ces variables, on a*

$$\mathbb{P}\left[(X_1 = x_1) \cap \dots \cap (X_n = x_n)\right] = \mathbb{P}(X_1 = x_1) \times \dots \times \mathbb{P}(X_n = x_n)$$

Remarque. *Le fait que les X_i soient deux à deux indépendantes ne suffit pas à ce que les variables $X_1, \dots, X_n$ soient indépendantes entre elles, au sens de la définition précédente.*

Par exemple, si l'on lance deux pièces équilibrées indépendamment l'une de l'autre, et que l'on note :

- $X_1 = \begin{cases} 1 & \text{si le premier lancer donne pile} \\ -1 & \text{sinon} \end{cases}$
- $X_2 = \begin{cases} 1 & \text{si le deuxième lancer donne pile} \\ -1 & \text{sinon} \end{cases}$
- $X_3 = X_1 \times X_2$

On a $X_1 \perp\!\!\!\perp X_2$, $X_1 \perp\!\!\!\perp X_3$ et $X_2 \perp\!\!\!\perp X_3$, car pour tous $i, k \in \{1; 2; 3\}$ tels que $i \neq k$,

$$\mathbb{P}_{(X_k = -1 \text{ ou } 1)}(X_i = -1 \text{ ou } 1) = \frac{1}{2}$$

Pourtant, X_1, X_2, X_3 ne sont pas indépendantes, car par exemple, la connaissance de deux d'entre elles détermine la troisième de façon déterministe.

Théorème. *Si $X_1, \dots, X_n$ sont indépendantes, alors :*

1. *Pour tout $A_1 \subset X_1(\Omega)$, ..., pour tout $A_n \subset X_n(\Omega)$, les évènements $(X_1 \in A_1)$, ..., $(X_n \in A_n)$ sont indépendants.*
2. *Pour toutes fonctions $f_1, \dots, f_n$, les variables aléatoires $f_1(X_1), \dots, f_n(X_n)$ sont indépendantes.*
3. *Pour toutes fonctions f et g, et pour tout $k \in [\![1; n]\!]$, alors les variables $f(X_1, \dots, X_k)$ et $g(X_{k+1}, \dots, X_n)$ sont indépendantes.*
4. *De façon plus générale, si $I_1, \dots, I_p$ sont des sous-ensembles disjoints de $[\![1; n]\!]$, et si :*
   - *Y_1 est une v.a. du type $Y_1 = f_1\left((X_i)_{i \in I_1}\right)$*
   - $\vdots$
   - *Y_p est une v.a. du type $Y_p = f_p\left((X_i)_{i \in I_p}\right)$*

   Alors : les v.a. $Y_1, \dots, Y_p$ sont indépendantes.

*Les deux derniers points sont généralement appelés **lemme des coalitions**, car ils expriment le fait que des coalitions disjointes de variables indépendantes sont indépendantes entre elles.*

Démonstration. Admise, car fastidieuse, et peu intéressante. □

5 Espérance

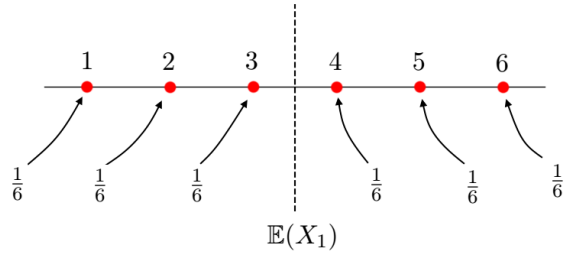
Définition. *Soit X une v.a réelle. On note $X(\Omega) = \{x_1, \dots, x_n\}$ et $p_i := \mathbb{P}(X = x_i)$.*
*On appelle **espérance de X**, et on note $\mathbb{E}(X)$, le nombre :*

$$\mathbb{E}(X) := \sum_{i=1}^{n} p_i x_i$$

Interprétation : moyenne des valeurs possibles de X, pondérée (c-à-d coefficientée) par les p_i.

Exemples :

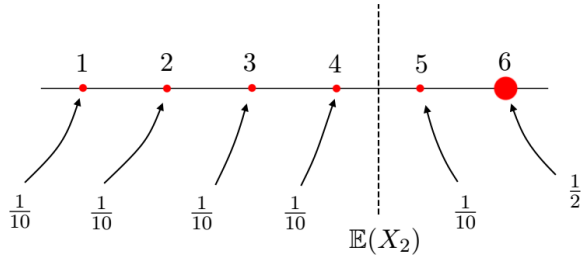
1. On note X_1 le résultat d'un dé équilibré.



$$\begin{aligned}\mathbb{E}(X_1) &= \left(\frac{1}{6} \times 1\right) + \left(\frac{1}{6} \times 2\right) + \dots + \left(\frac{1}{6} \times 6\right) \\ &= \frac{1}{6}\left(\frac{6 \times 7}{2}\right) \\ &= \boxed{3,5}\end{aligned}$$

2. On note X_2 le résultat d'un dé pipé tel que

$$\mathbb{P}(X_2 = 1) = \mathbb{P}(X_2 = 2) = \dots = \mathbb{P}(X_2 = 5) = \frac{1}{10} \qquad \text{et} \qquad \mathbb{P}(X_2 = 6) = \frac{1}{2}$$

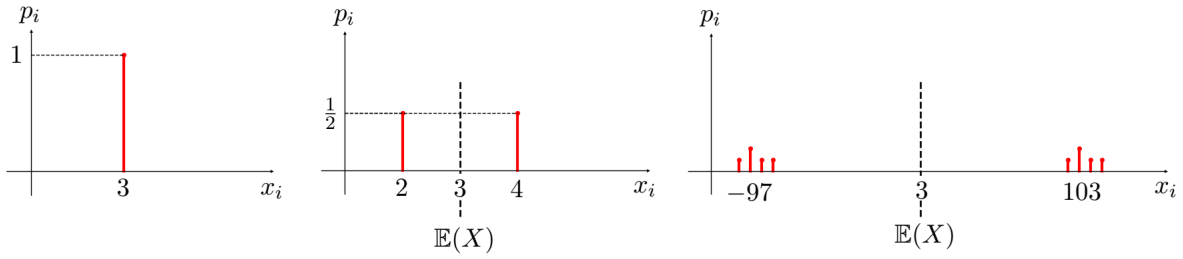


$$\begin{aligned}\mathbb{E}(X_2) &= \left(\frac{1}{10} \times 1\right) + \dots + \left(\frac{1}{10} \times 5\right) + \left(\frac{1}{2} \times 6\right) \\ &= \boxed{4,5}\end{aligned}$$

Remarque. *Si $x_1 < x_2 < \ldots < x_n$, alors*

$$x_1 \leqslant \underbrace{p_1x_1 + p_2x_2 + \cdots + p_nx_n}_{=\mathbb{E}(X)} \leqslant x_n$$

Intuition : $\mathbb{E}(X)$ est un indicateur de **localisation moyenne** de X. Voici présentées ci-dessous la loi de trois variables, toutes d'espérance égale à 3.



A elle seule, l'espérance de X est une information assez pauvre sur la loi de X. En particulier, ce n'est pas forcément une valeur typique de X.

Vocabulaire. *On dit que X est **centrée** si $\mathbb{E}(X) = 0$.*

Théorème (Formule de transfert)**.** *Dans les mêmes notations que la définition, si $f : X(\Omega) \to \mathbb{R}$, alors*

$$\mathbb{E}\left(f(X)\right) = \sum_{i=1}^{n} p_i f(x_i)$$

ou encore : $\mathbb{E}\left(f(X)\right) = \sum_{x \in X(\Omega)} \mathbb{P}(X = x) f(x)$

Exemples : $\mathbb{E}\left(e^X + 2X\right) = \sum_{i=1}^{n} p_i \left(e^{x_i} + 2x_i\right)$; et $\mathbb{E}\left(X^2\right) = \sum_{i=1}^{n} p_i x_i^2$.

Démonstration. Sur chaque évènement élémentaire $(X = x_i)$ du SCE adapté à X, $f(X)$ vaut non pas x_i mais $f(x_i)$. On en déduit :

Évè. du SCE $(X = x_i)_{i \in [\![1;n]\!]}$	$X = x_1$	$X = x_2$	$X = x_3$	$\ldots$	$\ldots$	$X = x_n$
Proba de l'évè.	p_1	p_2	p_3	$\ldots$	$\ldots$	p_n
Valeur de X	x_1	x_2	x_3	$\ldots$	$\ldots$	x_n
Valeur de $f(X)$	$f(x_1)$	$f(x_2)$	$f(x_3)$	$\ldots$	$\ldots$	$f(x_n)$

□

Cf exo 3

Remarque. *Appliquée à un couple de v.a. (X, Y), la formule de transfert donne*

$$\mathbb{E}\left(f(X,Y)\right) = \sum_{(x,y) \in (X,Y)(\Omega)} \mathbb{P}(X = x, Y = y) f(x,y)$$

Avec les notations du cours, on obtient la somme double

$$\mathbb{E}\left(f(X,Y)\right) = \sum_{i} \sum_{j} p_{i,j} f(x_i, y_j)$$

Plus généralement, pour un n-uplet de variables, on a

$$\begin{aligned}\mathbb{E}\left(f(X_1,\dots,X_n)\right) &= \sum_{(x_1,\dots,x_n)\in E_1\times\cdots\times E_n} \mathbb{P}\left(X_1=x_1,\dots,X_n=x_n\right) f(x_1,\dots,x_n)\\ &= \sum_{x_1\in E_1}\dots\sum_{x_n\in E_n} \mathbb{P}\left(X_1=x_1,\dots,X_n=x_n\right) f(x_1,\dots,x_n)\end{aligned}$$

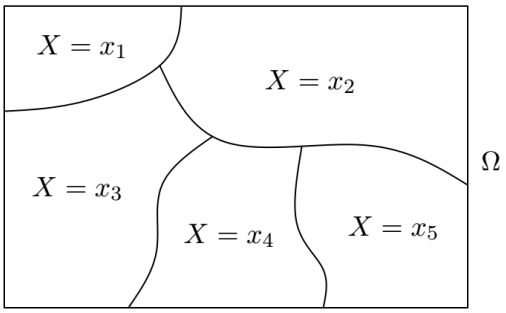
Remarque (Une autre définition de $\mathbb{E}(X)$).

On a défini $\mathbb{E}(X)$ comme suit :

$$\mathbb{E}(X)=\sum_{i=1}^{n} p_i x_i$$

ce que l'on peut interpréter comme une somme sur la partition de Ω :

$$\Omega=\bigcup_{i=1}^{n}(X=x_i)$$

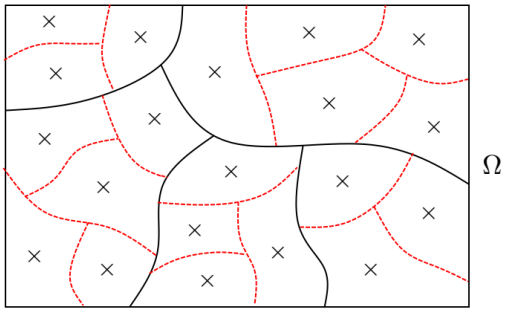


On peut aussi définir $\mathbb{E}(X)$ de la façon suivante :

$$\mathbb{E}(X)=\sum_{\omega\in\Omega}\mathbb{P}\left(\{\omega\}\right)X(\omega)$$

en sommant sur la partition élémentaire de Ω :

$$\Omega=\bigcup_{\omega\in\Omega}\{\omega\}$$



On passe de cette somme à la première par regroupement des ω selon leur valeur $X(\omega)$. Cette définition est peu utile en pratique, mais l'est davantage dans les démonstrations théoriques.

Proposition (Propriétés de l'espérance)**.** *Soit X et Y deux v.a.r.*

- ***Espérance d'une constante :*** $\forall\lambda\in\mathbb{R}$, $\mathbb{E}(\lambda)=\lambda$.
- ***Linéarité :*** $\forall\lambda,\mu\in\mathbb{R}$, $\mathbb{E}(\lambda X+\mu Y)=\lambda\mathbb{E}(X)+\mu\mathbb{E}(Y)$.
- ***Positivité :*** *Si $X\geqslant 0$ (c-à-d $\forall\omega\in\Omega$, $X(\omega)\geqslant 0$), alors* $\mathbb{E}(X)\geqslant 0$.
- ***Croissance :*** *Si $X\leqslant Y$ (c-à-d $\forall\omega\in\Omega$, $X(\omega)\leqslant Y(\omega)$), alors* $\mathbb{E}(X)\leqslant\mathbb{E}(Y)$.
  Inégalité triangulaire (conséquence de la croissance) : $|\mathbb{E}(X)|\leqslant\mathbb{E}(|X|)$.

Démonstration. On démontre la linéarité.

$$\begin{aligned}\mathbb{E}(\lambda X+\mu Y) &= \sum_{\omega\in\Omega}\mathbb{P}\left(\{\omega\}\right)\left(\lambda X(\omega)+\mu Y(\omega)\right)\\ &= \lambda\sum_{\omega\in\Omega}\mathbb{P}\left(\{\omega\}\right)X(\omega)+\mu\sum_{\omega\in\Omega}\mathbb{P}\left(\{\omega\}\right)Y(\omega)=\lambda\mathbb{E}(X)+\mu\mathbb{E}(Y)\end{aligned}$$

□

Exemple : Si $Y\sim\mathcal{U}\left(\llbracket 0;n\rrbracket\right)$, on avait calculé que $\mathbb{E}(Y)=\dfrac{n}{2}$.
On a alors $\mathbb{E}(2Y+1)=2\mathbb{E}(Y)+\mathbb{E}(1)=n+1$.

En revanche, pour $\mathbb{E}\left(2^Y\right)$, la formule du transfert est nécessaire.

Remarque. ⚠ *En général, $\mathbb{E}(XY)\neq\mathbb{E}(X)\cdot\mathbb{E}(X)$, exactement comme* $\displaystyle\int_a^b fg\neq\int_a^b f\times\int_a^b g$.

Théorème. *Si $X \perp\!\!\!\perp Y$, alors $\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y)$.*
Attention, la réciproque est fausse !

De façon générale, si $X_1, \dots, X_n$ sont indépendantes, alors on a

$$\mathbb{E}(X_1 \dots X_n) = \prod_{i=1}^{n} \mathbb{E}(X_i)$$

Corollaire. *Si $X \perp\!\!\!\perp Y$, alors pour toutes fonctions réelles f et g définies respectivement sur $X(\Omega)$ et $Y(\Omega)$:*

$$\mathbb{E}\left(f(X)g(Y)\right) = \mathbb{E}\left(f(X)\right)\,\mathbb{E}\left(g(Y)\right)$$

Démonstration. (du théorème)

► Si $X(\Omega) = \{x_1, \dots, x_n\}$ et $Y(\Omega) = \{y_1, \dots, y_p\}$, alors par définition, par la loi du couple (X, Y),

$$\begin{aligned}
\mathbb{E}(XY) &= \sum_{i=1}^{n}\sum_{j=1}^{p} \mathbb{P}\left((X = x_i) \cap (Y = y_j)\right) x_i y_j \\
&= \sum_{i=1}^{n}\sum_{j=1}^{p} \mathbb{P}(X = x_i) \times \mathbb{P}(Y = y_j) x_i y_j \\
&= \sum_{i=1}^{n} \mathbb{P}(X = x_i) x_i \underbrace{\sum_{j=1}^{p} \mathbb{P}(Y = y_j) y_j}_{=\mathbb{E}(Y)} \\
&= \mathbb{E}(Y) \underbrace{\sum_{i=1}^{n} \mathbb{P}(X = x_i) x_i}_{=\mathbb{E}(X)} \\
&= \mathbb{E}(X)\,\mathbb{E}(Y)
\end{aligned}$$

► Pour n v.a., il s'agit d'une récurrence sur n. L'hérédité découle du calcul suivant :

$$\begin{aligned}
\mathbb{E}\left(\underbrace{X_1 \times \dots \times X_n}_{\text{indép. de } X_{n+1}} \times X_{n+1}\right) &= \mathbb{E}\left(X_1 \dots X_n\right)\mathbb{E}(X_{n+1}) \\
&= \mathbb{E}(X_1) \dots \mathbb{E}(X_n)\mathbb{E}(X_{n+1}) \qquad \text{(par H.R.)}
\end{aligned}$$

□

Quatre exemples fondamentaux

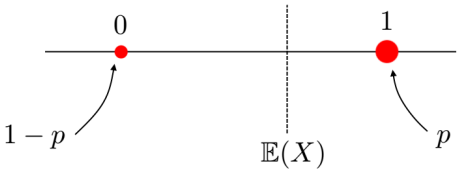
1. **Cas $\boldsymbol{X = \mathbb{1}_A}$, l'indicatrice d'un évènement A**
   Comme $X = \begin{cases} 1 & \text{si} \quad A \text{ se réalise} \\ 0 & \text{sinon} \end{cases}$, on a $X(\Omega) = \{0; 1\}$.
   Ainsi, $\mathbb{E}(X) = \mathbb{P}(X = 1) \times 1 + \mathbb{P}(X = 0) \times 0 = \mathbb{P}(A) \times 1 + \mathbb{P}\left(\overline{A}\right) \times 0$; donc $\mathbb{E}(X) = \mathbb{P}(A)$.

2. **Cas $\boldsymbol{X \sim \mathcal{U}\left(\{x_1, \dots, x_n\}\right)}$**
   On rappelle que $X(\Omega) = \{x_1, \dots, x_n\}$, et pour tout $i \in [\![1; n]\!]$, $\mathbb{P}(X = x_i) = \dfrac{1}{n}$.
   Donc, $\mathbb{E}(X) = \displaystyle\sum_{i=1}^{n} \frac{1}{n} x_i = \frac{1}{n}\sum_{i=1}^{n} x_i$. L'espérance de X est alors simplement la moyenne des x_i.

3. **Cas $X \sim \mathcal{B}(p)$**

$X(\Omega) = \{0,1\}$, et $\begin{cases} \mathbb{P}(X=1) = p \\ \mathbb{P}(X=0) = 1-p \end{cases}$.

$\mathbb{E}(X) = p \times 1 + (1-p) \times 0$, donc $\boxed{\mathbb{E}(X) = p}$.



4. **Cas $X \sim \mathcal{B}(n,p)$**

On a vu que $\mathcal{B}(n,p)$ est la loi de la somme de n v.a.r. $X_1, \dots, X_n$ indépendantes suivantes la loi $\mathcal{B}(p)$. Donc, par linéarité,

$$\mathbb{E}(X) = \mathbb{E}(X_1 + \cdots + X_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) = p + \cdots + p\,, \quad \text{donc :} \quad \boxed{\mathbb{E}(X) = np}$$

Cf exo 4

6 Variance et écart-type

Définition. *Soit X une v.a. réelle sur $(\Omega, \mathbb{P})$.*

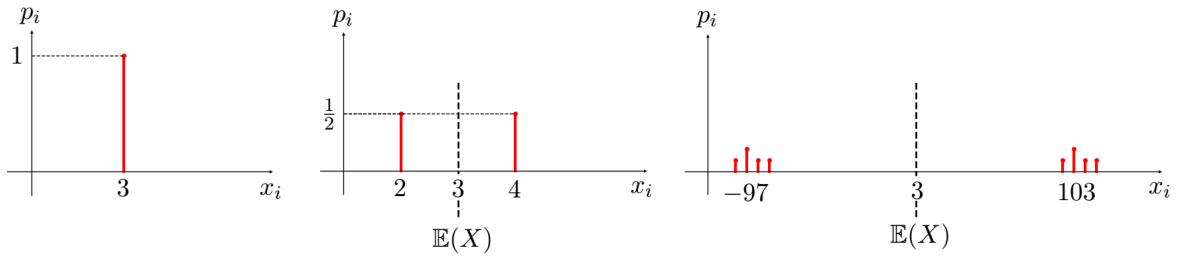
- *La **variance** de X est le nombre $\mathbb{V}(X) := \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$*
- *L'**écart-type** de X est le nombre $\sigma(X) := \sqrt{\mathbb{V}(X)}$.*

Remarque. *$\mathbb{V}(X)$ peut aussi être définie par $\mathbb{V}(X) = \mathbb{E}\left(X^2\right) - (\mathbb{E}(X))^2$.*
En effet, $(X - \mathbb{E}(X))^2 = X^2 - 2X\mathbb{E}(X) + (\mathbb{E}(X))^2$, donc :

$$\begin{aligned} \mathbb{V}(X) &= \mathbb{E}\left(X^2 - 2X\mathbb{E}(X) + (\mathbb{E}(X))^2\right) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)\,\mathbb{E}(X) + (\mathbb{E}(X))^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \end{aligned}$$

Interprétation : $|X - \mathbb{E}(X)|$ est la distance entre X et son espérance. Donc, $\mathbb{E}\left((X - \mathbb{E}(X))^2\right)$ peut être vu comme une moyenne des écarts quadratiques (c-à-d des écarts mis au carré) entre X et $\mathbb{E}(X)$.

La variance $\mathbb{V}(X)$ et l'écart-type $\sigma(X)$ sont des indicateurs de dispersion de la v.a. X autour de son espérance. Reprenons les trois exemples de lois qui avaient la même espérance.

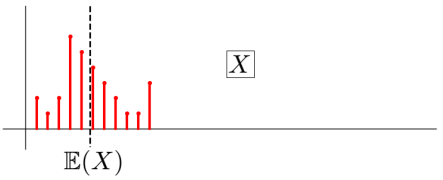


$\boxed{\mathbb{V}(X) = 0 \ ; \ \sigma(X) = 0}$ $\boxed{\mathbb{V}(X) = 1 \ ; \ \sigma(X) = 1}$ $\boxed{\mathbb{V}(X) \simeq 10\,000 \ ; \ \sigma(X) \simeq 100}$

Vocabulaire. *Une v.a.r. est dite **réduite** si sa variance vaut 1.*

Propriétés. *Soit X, Y deux v.a.r. et soit $\lambda \in \mathbb{R}$.*

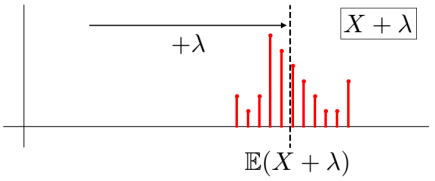
1. $\mathbb{V}(X+\lambda) = \mathbb{V}(X)$ *(Invariance par translation)*
2. $\mathbb{V}(\lambda X) = \lambda^2 \mathbb{V}(X)$
3. *Si $X \perp\!\!\!\perp Y$, alors* $\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y)$.



Démonstration.

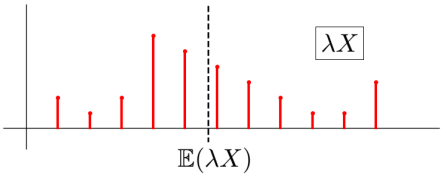
1. On a $(X+\lambda) - \mathbb{E}(X+\lambda) = X + \lambda - \mathbb{E}(X) - \lambda$. Ainsi,

$$\begin{aligned}\mathbb{E}\left(((X+\lambda) - \mathbb{E}(X+\lambda))^2\right) &= \mathbb{E}\left((X - \mathbb{E}(X))^2\right) \\ &= \mathbb{V}(X)\end{aligned}$$



2. On a $(\lambda X) - \mathbb{E}(\lambda X) = \lambda\,(X - \mathbb{E}(X))$. Ainsi,

$$\begin{aligned}\mathbb{E}\left((\lambda X - \mathbb{E}(\lambda X))^2\right) &= \mathbb{E}\left(\lambda^2\,(X - \mathbb{E}(X))^2\right) \\ &= \lambda^2\,\mathbb{E}\left((X - \mathbb{E}(X))^2\right) \\ &= \lambda^2 \mathbb{V}(X)\end{aligned}$$



3. (voir prochaine partie sur la covariance)

□

Remarque. *Si $\sigma(X) > 0$, alors la variable* $\dfrac{X - \mathbb{E}(X)}{\sigma(X)}$ *est centrée réduite.*

En effet, par linéarité de l'espérance,

$$\mathbb{E}\left(\frac{X - \mathbb{E}(X)}{\sigma(X)}\right) = \frac{1}{\sigma(X)}\left(\mathbb{E}(X) - \mathbb{E}(X)\right) = 0$$

Et, par les propriétés de la variance vues ci-dessus,

$$\mathbb{V}\left(\frac{X - \mathbb{E}(X)}{\sigma(X)}\right) = \frac{1}{(\sigma(X))^2}\mathbb{V}\left(X - \mathbb{E}(X)\right) = \frac{1}{(\sigma(X))^2}\mathbb{V}(X) = \frac{(\sigma(X))^2}{(\sigma(X))^2} = 1$$

Trois exemples fondamentaux

1. **Cas $\boldsymbol{X = \lambda \in \mathbb{R}}$ (constante)**. Dans ce cas :

$$\mathbb{V}(\boldsymbol{X}) = \boldsymbol{0}$$

En effet, $X - \mathbb{E}(X) = \lambda - \lambda = 0$, donc $\mathbb{E}\left((X - \mathbb{E}(X))^2\right) = 0$.

2. **Cas $\boldsymbol{X \sim \mathcal{B}(p)}$**. On a alors :

$$\mathbb{V}(\boldsymbol{X}) = \boldsymbol{p(1-p)}$$

En effet, $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$, et on sait que $\mathbb{E}(X) = p$. Par ailleurs, $\mathbb{E}(X^2) = p \times 1^2 + (1-p) \times 0^2 = p$. Ainsi, $\mathbb{V}(X) = p - p^2 = p(1-p)$.

3. **Cas $\boldsymbol{X \sim \mathcal{B}(n,p)}$** . Dans ce cas :

$$\mathbb{V}(\boldsymbol{X}) = \boldsymbol{np(1-p)}$$

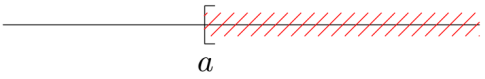
En effet, on sait que si $X_1, \ldots, X_n$ sont des v.a. indépendantes suivant la même loi de Bernoulli de paramètre p, alors $X := X_1 + \cdots + X_n \sim \mathcal{B}(n,p)$. Ainsi, par indépendance, la variance de la somme est la somme des variances, d'où :

$$\mathbb{V}(X) = \sum_{i=1}^{n} \mathbb{V}(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

Théorème. *Soit X une v.a.r.*

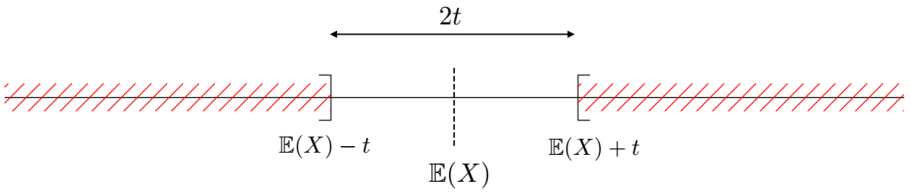
1. **(Inégalité de Markov)** *Si X est positive, alors :*

$$\forall a > 0, \quad \mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{E}(X)}{a}$$



2. **(Inégalité de Bienaymé-Tchebychev)** *On a :*

$$\forall t > 0, \quad \mathbb{P}\left(|X - \mathbb{E}(X)| \geqslant t\right) \leqslant \frac{\mathbb{V}(X)}{t^2}$$



Démonstration.

1. Soit $a > 0$. On a

$$\begin{aligned}
\mathbb{E}(X) = \sum_{k \in X(\Omega)} k\,\mathbb{P}(X = k) &= \sum_{k<a} k\,\mathbb{P}(X = k) + \sum_{k \geqslant a} k\,\mathbb{P}(X = k) \\
&\geqslant \sum_{k \geqslant a} k\,\mathbb{P}(X = k) \qquad (\text{car } k \geqslant 0) \\
&\geqslant a \sum_{k \geqslant a} \mathbb{P}(X = k) = a\mathbb{P}(X \geqslant a)
\end{aligned}$$

   On obtient bien

$$\mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{E}(X)}{a}$$

2. Soit $t > 0$.

$$\begin{aligned}
\mathbb{P}\left(|X - \mathbb{E}(X)| \geqslant t\right) &= \mathbb{P}\left(|X - \mathbb{E}(X)|^2 \geqslant t^2\right) \\
&\leqslant \frac{\mathbb{E}\left((X - \mathbb{E}(X))^2\right)}{t^2} \qquad (\text{par Markov}) \\
&= \frac{\mathbb{V}(X)}{t^2}
\end{aligned}$$

□

Exemple fondamental : Si $X_1, \dots, X_n$ sont n v.a. de même loi, d'espérance m et d'écart-type $\sigma > 0$, la moyenne des X_i :

$$M_n := \frac{1}{n}\sum_{i=1}^{n} X_i$$

a une espérance

$$\mathbb{E}(M_n) = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\mathbb{E}(X_i)}_{=m} = \frac{1}{n} \times nm = m$$

et une variance (par indépendance)

$$\mathbb{V}(M_n) = \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{\mathbb{V}(X_i)}_{=\sigma^2} = \frac{1}{n^2} \times n\sigma^2 = \frac{\sigma^2}{n}$$

Lorsque $n \to +\infty$, cette variance tend vers 0. On en déduit donc que lorsque $n \to +\infty$, la moyenne M_n a sa loi qui se concentre autour de l'espérance théorique m. Ce principe s'appelle la loi des grands nombres. Il s'observe très souvent dans la vie par le fait qu'une moyenne faite sur de nombreuses occurrences donne une estimation fiable, alors qu'une moyenne sur un petit nombre d'expériences donne une moyenne hasardeuse.

L'inégalité de Bienaymé-Tchebychev permet alors d'obtenir ce qu'on appelle une inégalité de concentration, c'est-à-dire un contrôle de la probabilité d'être éloigné de son espérance :

$$\forall t > 0\,,\ \mathbb{P}\left(|M_n - m| \geqslant t\right) \leqslant \frac{\mathbb{V}(M_n)}{t^2} = \frac{\sigma^2}{nt^2}$$

Cf exo 10

7 Covariance

Définition. *Soit X, Y deux v.a.r. sur $(\Omega, \mathbb{P})$.*
*On appelle **covariance** de X et Y le réel*

$$Cov(X,Y) = \mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right)$$

Remarques.

- *C'est l'espérance du produit des v.a. centrées associées à X et Y.*
- $Cov(X,X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{V}(X)$.
- *Intuitivement, le signe de $Cov(X,Y)$ indique si les valeurs élevées de l'une coïncident statistiquement avec des valeurs élevées de l'autre, ou bien au contraire à des valeurs faibles de l'autre.*
  - *Si $Cov(X,Y) > 0$ (imaginons même qu'il s'agisse d'une valeur très grande), cela signifie que le produit $\big(X - \mathbb{E}(X)\big)\big(Y - \mathbb{E}(Y)\big)$ est très souvent positif. Autrement dit, les quantités $X - \mathbb{E}(X)$ et $Y - \mathbb{E}(Y)$ ont souvent même signe. On comprend donc que X est supérieure à son espérance lorsque Y est elle-même supérieure à son espérance, et inversement. On dit que X et Y sont **corrélées positivement**. C'est le cas par exemple de la variable X égale au nombre de lunettes de soleil vendues et de Y le nombre de coups de soleil attrapés. Ces deux variables sont en effet hautes simultanément quand il fait beau, et basses simultanément lorsqu'il fait gris.*
  - *Dans le cas où $Cov(X,Y) < 0$, l'évènement $X \geqslant \mathbb{E}(X)$ coïncide souvent avec $Y \leqslant \mathbb{E}(Y)$ (et inversement), et on dit que X et Y sont **corrélées négativement**. C'est le cas par exemple de la variable X égale au nombre de lunettes de soleil vendues et de Y le nombre de personnes qui portent une écharpe.*
  - *Si $Cov(X,Y) = 0$, on dit que X et Y sont **décorrélées**. C'est par exemple le cas lorsque X et Y sont indépendantes (mais pas seulement), comme on va le voir.*

Les calculs de covariance ne démontrent en rien des causalités du type "X a un effet sur Y" (ce n'est pas le fait d'acheter des lunettes de soleil qui donne des coups de soleil), il mettent simplement en évidence une certaine simultanéité probabiliste.

Proposition. *Soit X, Y deux v.a.r. sur $(\Omega, \mathbb{P})$.*

$$Cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Démonstration. On développe :

$$(X - \mathbb{E}(X))(Y - \mathbb{E}(Y)) = XY - \mathbb{E}(X)Y - X\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y)$$

On en déduit, par linéarité de l'espérance,

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y)\\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\end{aligned}$$

□

Vocabulaire. *Si X et Y sont deux v.a.r. telles que $Cov(X,Y) = 0$, on dit que X et Y sont **décorrélées**.*

Corollaire. *Si X et Y sont indépendantes, alors elles sont décorrélées.*

Démonstration. Supposons $X \perp\!\!\!\perp Y$. On a alors $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$; cela implique $\mathrm{Cov}(X,Y) = 0$. □

Remarque. *Deux variables peuvent très bien être décorrélées mais ne pourtant pas être indépendantes. Par exemple, si on tire une boule dans une urne contenant 3 boules et qu'on définit*

$$X = \begin{cases} 1 & \text{si on tire la boule 1} \\ -1 & \text{si on tire la boule 2} \\ 0 & \text{si on tire la boule 3} \end{cases} \qquad \text{et} \qquad Y = \begin{cases} 1 & \text{si on tire la boule 1} \\ 1 & \text{si on tire la boule 2} \\ -2 & \text{si on tire la boule 3} \end{cases}$$

alors $\mathbb{E}(X) = \mathbb{E}(Y) = 0$ et $\mathbb{E}(XY) = \frac{1}{3}(1 - 1 + 0) = 0$, donc $Cov(X,Y) = 0$. Pourtant, elles ne sont pas indépendantes, car par exemple :

$$\mathbb{P}_{X=0}(Y = -2) = 1 \neq \mathbb{P}(Y = -2) = \frac{1}{3}$$

Proposition. *Soit X, Y, Z trois variables aléatoires à valeurs réelles, et $a, b \in \mathbb{R}$. On a :*

- ***Invariance par translation :*** $Cov(X + a, Y + b) = Cov(X,Y)$
- **Covariance par rapport à une constante :** $Cov(a, X) = 0$
- **Symétrie :** $Cov(X,Y) = Cov(Y,X)$
- **Bilinéarité :** $Cov(aX + bY, Z) = a\,Cov(X,Z) + b\,Cov(Y,Z)$

Démonstration. Cf exo 12 □

Théorème. *Si X, Y sont deux v.a.r. sur Ω, on a :*

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2Cov(X,Y)$$

*et donc si X et Y sont **indépendantes** :*

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$$

*Plus généralement, si $X_1, \ldots, X_n$ sont **indépendantes** :*

$$\mathbb{V}(X_1 + \cdots + X_n) = \mathbb{V}(X_1) + \cdots + \mathbb{V}(X_n)$$

Démonstration.

- Par bilinéarité, puis par symétrie,

$$\begin{aligned}\mathbb{V}(X+Y) &= \mathrm{Cov}(X+Y, X+Y) = \mathrm{Cov}(X,X) + \mathrm{Cov}(X,Y) + \mathrm{Cov}(Y,X) + \mathrm{Cov}(Y,Y) \\ &= \mathbb{V}(X) + \mathbb{V}(Y) + 2\mathrm{Cov}(X,Y)\end{aligned}$$

- Si $X \perp\!\!\!\perp Y$, on a $\mathrm{Cov}(X,Y) = 0$, et donc, $\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y)$.
- Dans le cas de n variables indépendantes, il s'agit d'une récurrence sur n, avec dans l'hérédité :

$$\begin{aligned}\mathbb{V}(\underbrace{X_1 + \cdots + X_n}_{\text{indep de } X_{n+1}} + X_{n+1}) &= \mathbb{V}(X_1 + \cdots + X_n) + \mathbb{V}(X_{n+1}) \\ &\underset{\text{HR}}{=} \mathbb{V}(X_1) + \cdots + \mathbb{V}(X_n) + \mathbb{V}(X_{n+1})\end{aligned}$$

□

Annexe - une démonstration intéressante de la formule de la variance d'une binomiale

Démonstration. Supposons que $X \sim \mathcal{B}(n,p)$. Posons $q := 1-p$; on a donc $p+q=1$.

On veut utiliser la formule $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

Or, par définition,

$$\mathbb{E}(X) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \times k$$

et par la formule de transfert,

$$\mathbb{E}(X^2) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \times k^2$$

Idée futée : Pour tout $x \in \mathbb{R}$, posons $f(\boldsymbol{x}) := \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}\, \boldsymbol{x^k} = (p\,\boldsymbol{x}+q)^n$, par la formule du binôme.

La fonction f est une fonction polynomiale de degré n, donc elle est dérivable sur $\mathbb{R}$. Ainsi, pour tout $x \in \mathbb{R}$,

$$f'(\boldsymbol{x}) = \sum_{k=1}^{n} \binom{n}{k} p^k q^{n-k} k\, \boldsymbol{x^{k-1}} = np(p\,\boldsymbol{x}+q)^{n-1}$$

$$f''(\boldsymbol{x}) = \sum_{k=2}^{n} \binom{n}{k} p^k q^{n-k} k(k-1)\, \boldsymbol{x^{k-2}} = n(n-1)p^2(p\,\boldsymbol{x}+q)^{n-2}$$

On a alors :

- $\mathbb{E}(X) = f'(1) = np \times (p+q)^{n-1} = np$
- Comme $k^2 = k(k-1)+k$, on a $\mathbb{E}(X^2) = f''(1) + f'(1) = n(n-1)p^2 + np$

Finalement,

$$\mathbb{V}(X) = n(n-1)p^2 + np - (np)^2 = np\,((n-1)p + 1 - np) = np(1-p)$$

□

Schéma récapitulatif de cette technique : on part de la formule du binôme, puis on dérive successivement à gauche et à droite.

$$\begin{array}{ccc} \displaystyle\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} x^k & = & (px+q)^n \\ \downarrow & & \downarrow \\ \displaystyle\sum_{k=1}^{n} \binom{n}{k} p^k q^{n-k} k x^{k-1} & = & np(px+q)^{n-1} \\ \downarrow & & \downarrow \\ \displaystyle\sum_{k=2}^{n} \binom{n}{k} p^k q^{n-k} k(k-1) x^{k-2} & = & n(n-1)p^2(px+q)^{n-2} \end{array}$$